# why open and machine readable?

cheap data changes the epistemology.

$$OPS = AB(H + BB + HBP) + TB(AB + BB + SF + HBP) / AB(AB + BB + SF + HBP)$$

opening data attempts to reflect the epistemology.

# Open-source software

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed.
*(February 2010)*

**Open-source software** (**OSS**) is computer software with its source code made available and licensed with a license in which the copyright holder provides the rights to study, change and distribute the software at no cost to anyone and for any purpose. Open-source software is very often developed in a public, collaborative manner. Open-source software is the most prominent example of open-source development and often compared to (technically defined) user-generated content or (legally defined) open-content movements.[1]

A report by the Standish Group (from 2008) states that adoption of open-source software models has resulted in savings of about $60 billion per year to consumers.[2][3]

**Contents** [hide]

1 History

# 1. big data as we experience it.

# Your Kindle Library

View: Books OR Search your library GO

**Showing 1 - 15 of 112 items**

| Title | | |
|---|---|---|
| Predictably Irrational, R<br>Hidden Forces That Sha | | 7, 2013 |
| Someone Could Get Hu<br>Century Parenthood | | 6, 2013 |
| The Last Utopia: Humar | | 0, 2013 |
| Intuition Pumps And Ot | | 30, 2013 |
| Cooked: A Natural Histo | | 22, 2013 |
| The 22 Immutable Laws<br>Explained by the World' | | |
| The Ultimate Question 2<br>Edition): How Net Prom<br>Customer-Driven World | | |
| The Fourth Part of the V<br>Earth, and the Epic Sto<br>Its Name | | |
| The Player of Games (O | | |
| No Place to Hide | O'Harrow, Robert | March 14, 2013 |
| Against a Dark Background | Banks, Iain M. | March 7, 2013 |
| To Save Everything, Click Here: The Folly of Technological | Morozov, Evgeny | March 7, 2013 |

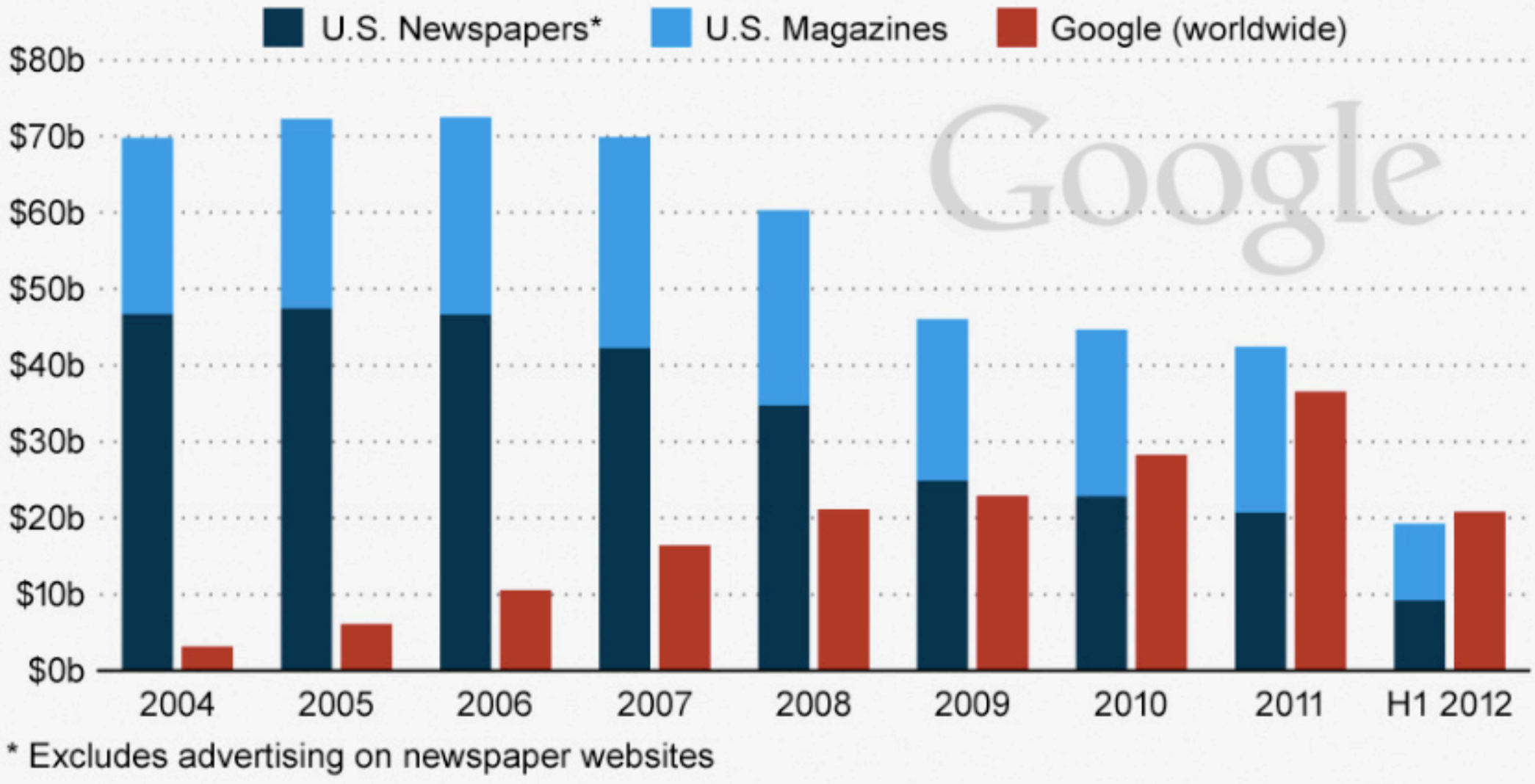
- Read Now
- Deliver to my...
- Download & transfer via USB
- Clear furthest page read...
- Delete from library

Actions...

Actions...

# Google Rakes In More Ad Dollars Than U.S. Print Media

Advertising revenue in billion U.S. dollars




* Excludes advertising on newspaper websites

statista The Statistics Portal

creative commons

Source: Google, NAA, PIB

## Market share of the main internet portals

**Share of global internet ad expenditure (%)**

| | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|
| Google | 34.9 | 40.3 | 42.5 | 41.9 | 44.1 |
| Microsoft | 8.1 | 7.9 | 4.2 | 4.0 | 4.0 |
| Yahoo! | 18.7 | 14.9 | 11.7 | 9.6 | 8.3 |
| AOL | 6.3 | 5.5 | 4.2 | 2.2 | 1.5 |
| Facebook | 0.2 | 0.4 | 0.6 | 1.4 | 3.1 |
| **Total** | **68.1** | **68.9** | **63.2** | **59.2** | **61.0** |

Overall the internet represents only 16 percent of global ad revenue according to ZenithOptimedia. TV, by comparison, is 40.2 percent of all ad expenditures. Hence Google's interest in building YouTube into a *bona fide* TV/cable alternative.

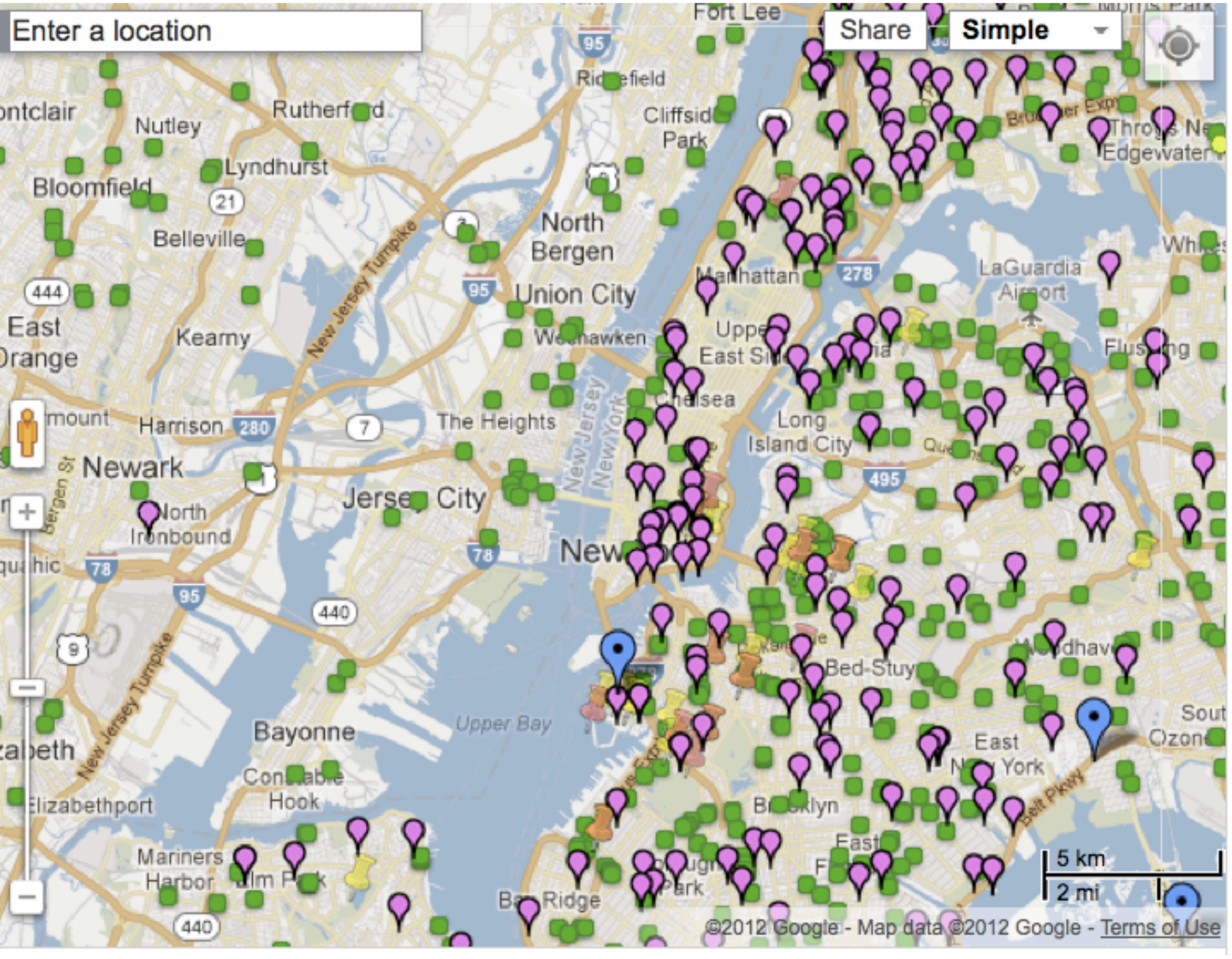
## Advertising expenditure by medium

***US$ million, current prices*** *Currency conversion at 2010 average rates.*

| | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|
| Newspapers | 94,600 | 91,495 | 89,868 | 88,785 | 88,446 |
| Magazines | 43,741 | 43,122 | 42,681 | 42,464 | 42,186 |
| Television | 176,627 | 184,290 | 193,735 | 203,608 | 215,737 |
| Radio | 32,017 | 32,903 | 33,667 | 34,827 | 35,923 |
| Cinema | 2,313 | 2,442 | 2,564 | 2,732 | 2,916 |
| Outdoor | 29,824 | 31,291 | 32,928 | 34,559 | 36,350 |
| Internet | 63,979 | 72,842 | 84,267 | 97,764 | 113,281 |
| **Total *** | **443,100** | **458,385** | **479,710** | **504,738** | **534,839** |

*Source: ZenithOptimedia*

# “open” data for “good”




## Superstorm Sandy: NYC

Sandy was a major storm that caused extensive flooding, physical damage, power and transportation outages, and fatalities in NY and much of the US East coast.

- Latest alerts, NYC Emergency Mgmt
- City of New York on Twitter
- NYC Emergency Mgmt on Facebook
- WNYC Transit Tracker
- NYC.gov information including volunteering and school status

How are you using this map? *Tell us.*

Reset to default view

☐ Pharmacies, restaurants, hotels

☑ Gas stations

Legend

- Gas available
- Gas inventory low or uncertain
- Gas not available
- Gas not available (no power)
- Gas not available (no power, no gas)

**NOTE:** It is not possible to predict the rate at

# 2. how might we assess the economic value of data?

**Browse By Category**

- Astronomy
- Biology
- Chemistry
- Climate
- Economics
- Encyclopedic
- Geographic
- Mathematics

**Developer Resources**

- Amazon Machine Images (AMIs)
- Articles & Tutorials
- Customer Apps

# Public Data Sets

Public Data Sets on AWS provides a centralized repository of public data sets that can be seamlessly integrated into AWS cloud-based applications. AWS is hosting the public data sets at no charge for the community, and like all AWS services, users pay only for the compute and storage they use for their own applications. Learn more about Public Data Sets on AWS and visit the Public Data Sets forum.

## Featured Public Data Sets

### 1000 Genomes Project

The 1000 Genomes Project, initiated in 2008, is an international public-private consortium that aims to build the most detailed map of human genetic variation available.

### Common Crawl Corpus

A corpus of web crawl data composed of 5 billion web pages. This data set is freely available on Amazon S3 and formatted in the ARC (.arc) file format.

### Google Books Ngrams

A data set containing Google Books n-gram corpuses. This data set is freely available on Amazon S3 in a Hadoop friendly file format and is licensed under a Creative Commons Attribution 3.0 Unported License. The original dataset is available from http://books.google.com/ngrams/.

commitment.

DATA.GOV
EMPOWERING PEOPLE
Search Data.gov
SEARCH
HOME
ABOUT
DATA
APPS
COMMUNITIES
METRICS
OPEN GOVERNMENT
BLOGS
VISUALIZED DATA HELPS
TRACK
HURRICANE
SANDY
2012-10-29
8:00 PM Mon EDT
92mph
2012-10-29
11:00 AM Mon EDT
92mph
SANDY
Latest Datasets
2010 Small Area Health Insurance...
2010 Small Area Health Insurance...
E3: Economy+Energy+Environment...
FY2011 Compensation and Pension by...
Gravesite locations of Veterans and...
Gravesite locations of Veterans and...
Gravesite locations of Veterans and...
Gravesite locations of Veterans and...
Gravesite locations of Veterans and...
Gravesite locations of Veterans and...
DATA AND APPS
COMMUNITIES
OPEN GOVERNMENT

commitment.

A peer-reviewed open-access journal

# Biodiversity Data Journal

Making your data count!

Open Access

## Biodiversity Data Journal

***Making your data count!***

"Science is a combination of gathering facts and making theories; neither can progress on its own. [...] In the history of science, the laborious accumulation of facts is the dominant mode, not a novelty."

Peter Norvig, Director of Research @Google Inc.

**Biodiversity Data Journal (BDJ)** is a community peer-reviewed, open-access, comprehensive online platform, designed to accelerate publishing, dissemination and sharing of biodiversity-related data of any kind. All structural elements of the articles – text, morphological descriptions, occurrences, data tables, etc. – will be treated and stored as DATA.

The journal will publish papers in biodiversity science containing taxonomic, floristic/faunistic, morphological, genomic, phylogenetic, ecological or environmental data on any taxon of any geological age from any part of the world with **no lower or upper** limit to manuscript size.



citation.

Home | Members | FAQs | Services | Resources | Events | Contact us

DataCite

Helping you to find, access, and reuse data

## DataCite Business Models Principles are online

Published by Frauke Ziedorn on 31 October 2012 - 1:08pm

DataCite's Business Models Principles document is now available for download. Intended for DataCite members and their clients, the document looks at DataCite member and client responsibilities, as well as best practices.

Read more

## DataCite Services

Published by Frauke Ziedorn on 18 October 2012 - 1:32pm

DataCite offers various services supporting the DOI registration and discovery of research data. These services include the Metadata Store, a statistics portal, the OAI-PMH Provider and many more. Please have a look at our Services website for more information!

Read more

Why cite data?

What is DataCite?

What do we do?

**DOI resolver**

Resolve a DOI string (e.g.

citation.

# Data Marketplace

With thousands of public and proprietary data sets, our data marketplace is a great resource of experimental and sample data for data scientists and developers.

## Geo APIs

Add geo-context and millions of locations to your apps and analytics, with data including Foursquare check-ins, ACS Census demographics, geo-Wikipedia articles, and more.

Learn more

## Social APIs

Power your websites and applications with industry-leading Twitter influence metrics, analytics, and raw data — built on our corpus of over 10 billion historical tweets.

Learn more

### Top Tags

locations list business geo location freebase places phone stores hours business-information

phone-numbers place-names store-hours aggdata population census government demographics

america health maps state care toxic facilities geonames facility release school community

More...

price.

## price

FREE (68)

PAID (88)

FREE TRIAL (32)

## category

AUTOMOTIVE, INDUSTRIAL, & AEROSPACE (3)

BUSINESS AND FINANCE (48) ▶

COMMUNICATIONS (5) ▶

CONSUMER GOODS & DISTRIBUTION (5)

DATA QUALITY SERVICES (12)

DEMOGRAPHICS (25)

DEVELOPER SERVICES (5)

GOVERNMENT (16)

HEALTH AND WELLNESS (10)

HIGH TECH & ELECTRONICS (3)

MANUFACTURING (2)

MEDIA & ENTERTAINMENT (2)

REAL ESTATE (15)

**143** Results in: DATA

Sort By: Date Added | Name | Publisher



### Worldwide Historical Weather Data

published by: Weather Trends International

Historical daily maximum temperature, minimum temperature, precipitation, dewpoint, sea level pressure, windspeed, and wind gust for thousands of locations around the world.

### Drug Prescribing by GP Practices in England

published by: Custom Web Apps, Ltd

Covering all general practices in England, the data includes figures on the number of prescription items that are dispensed each month and information relating to costs. For the first time, users can query the API using a combination of location (UK postcode) and ANY drug BNF code! For each GP practice, the total number of items that were prescribed and then dispensed, the total Net Ingredient Cost and the total Actual Cost of these items is shown.

### Barcelona Facilities

price.

“capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences”

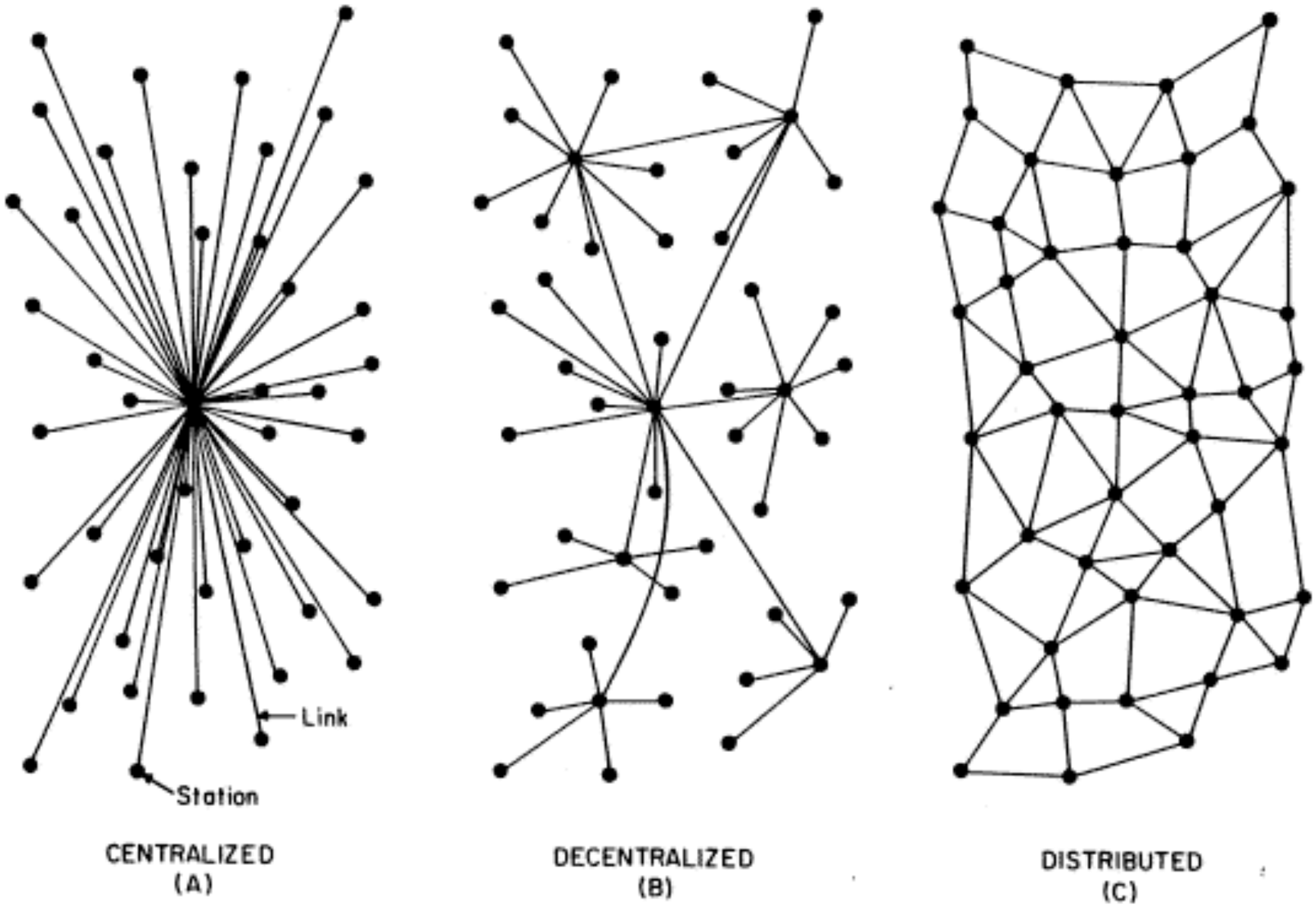
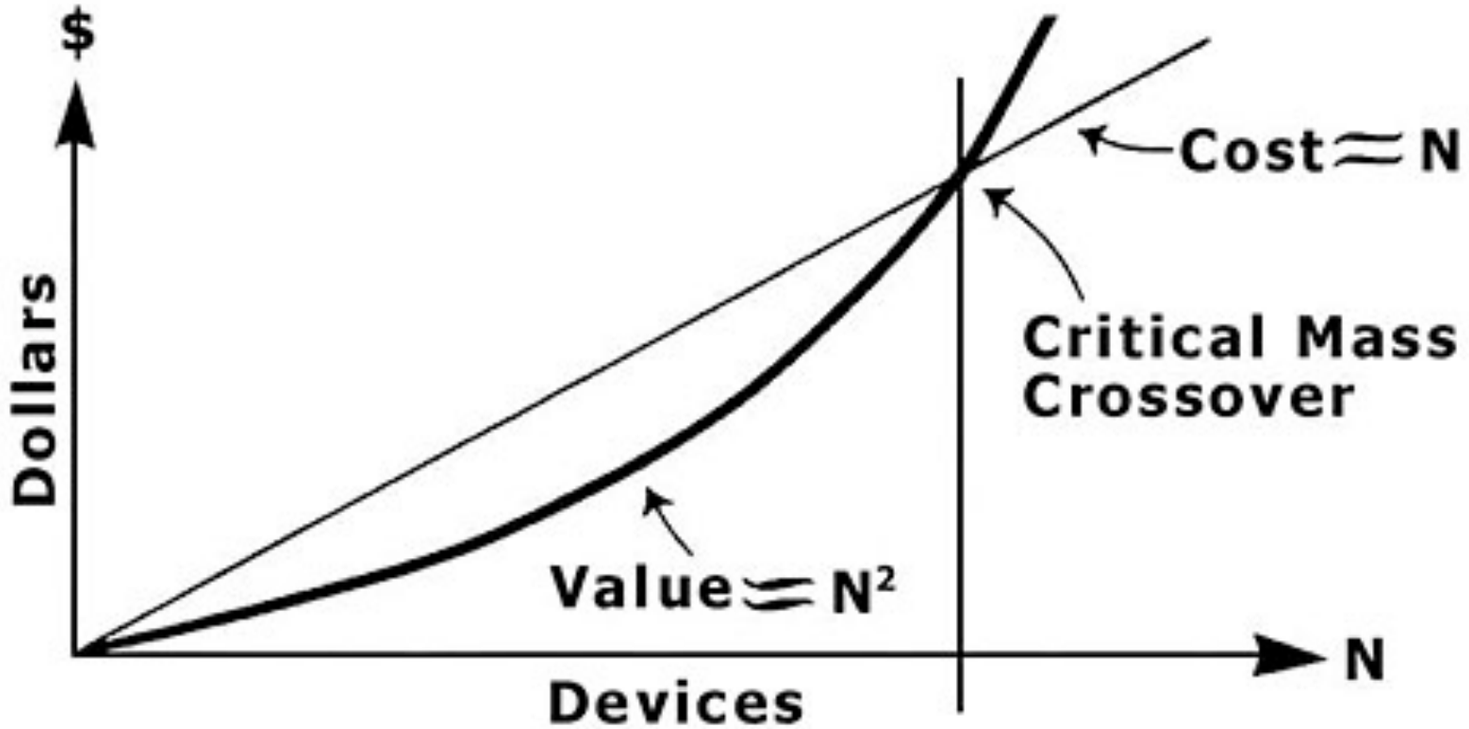
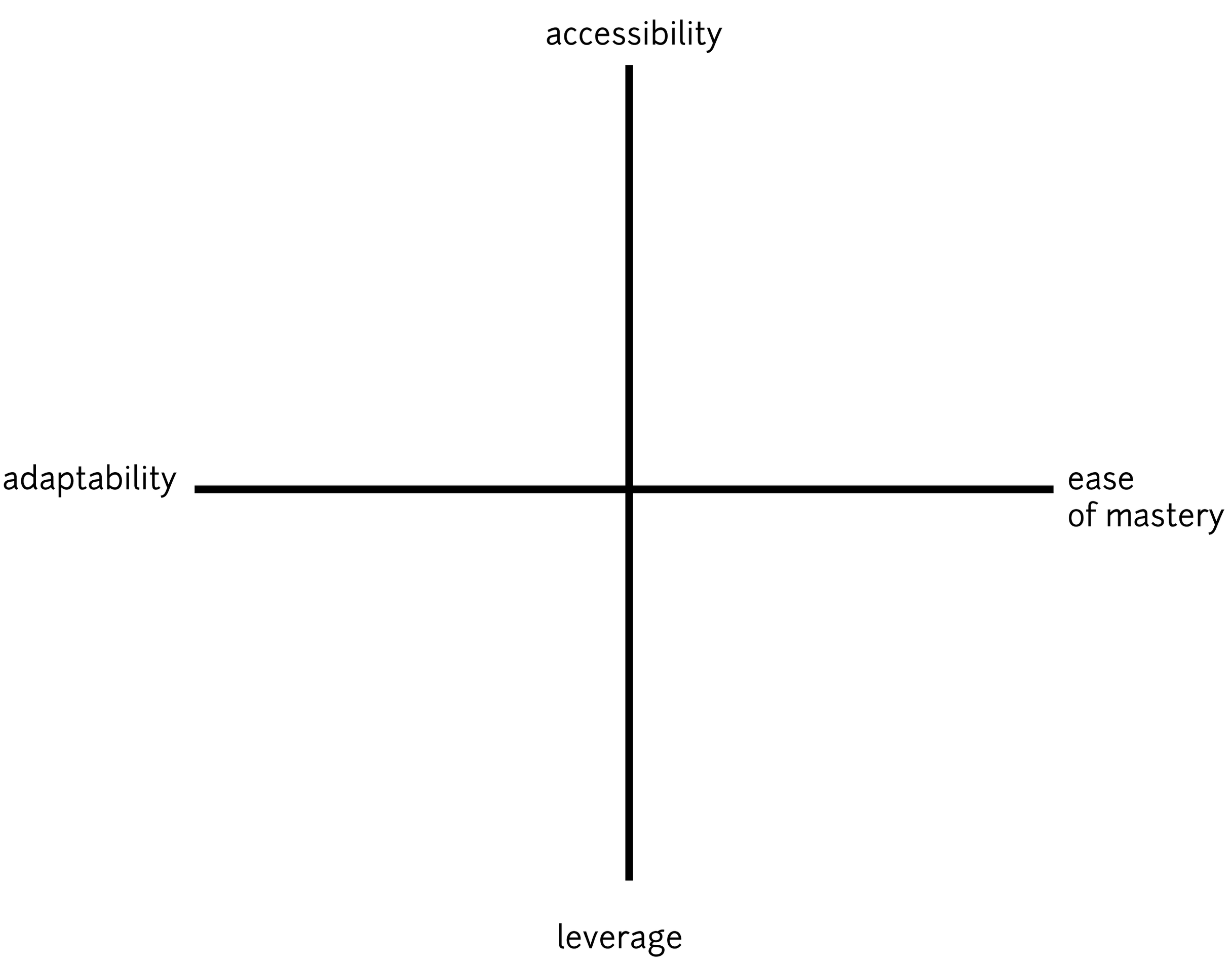



FIG. 1 — Centralized, Decentralized and Distributed Networks

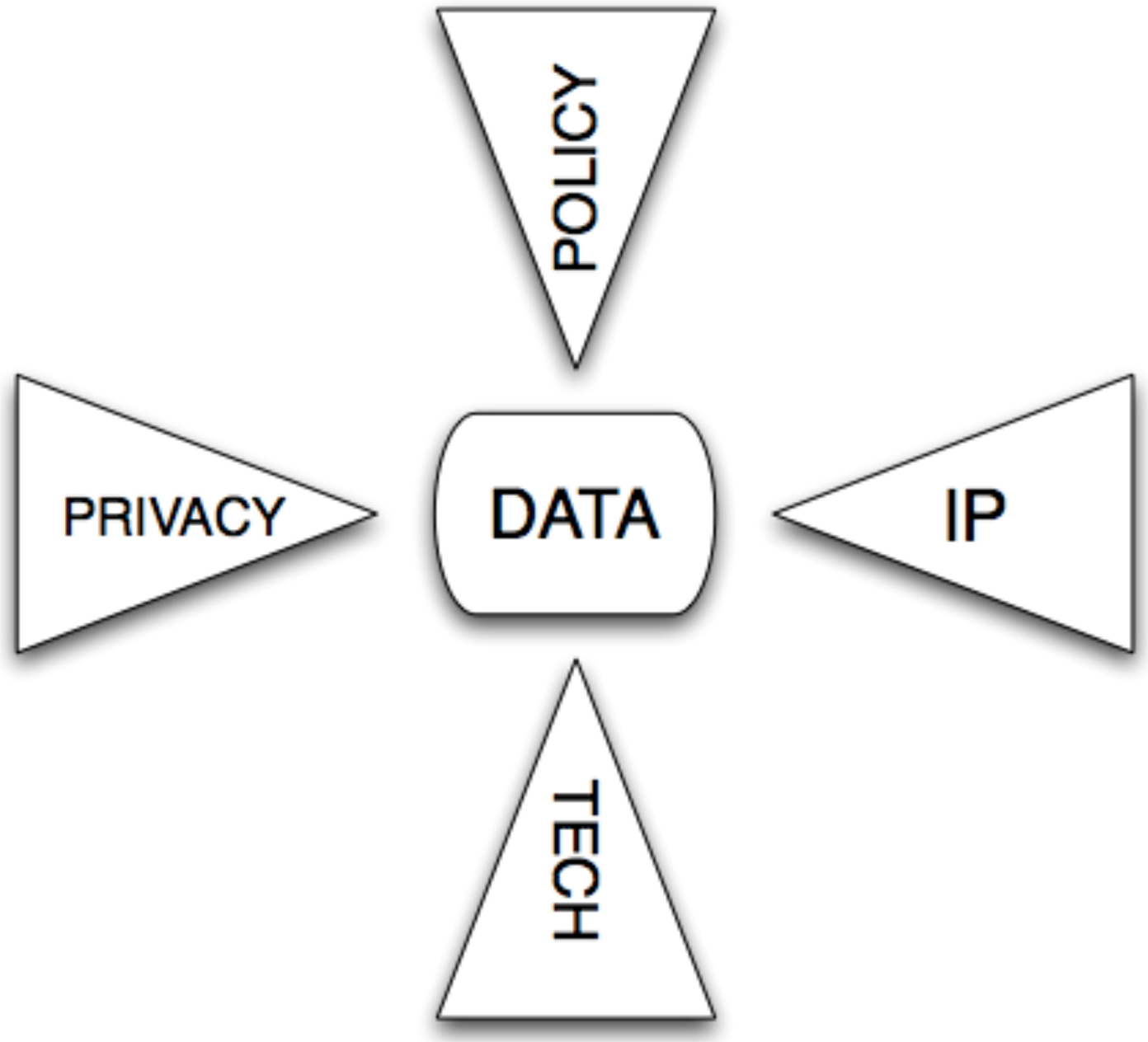
**The Systemic Value of Compatibly Communicating Devices Grows as the Square of Their Number:**

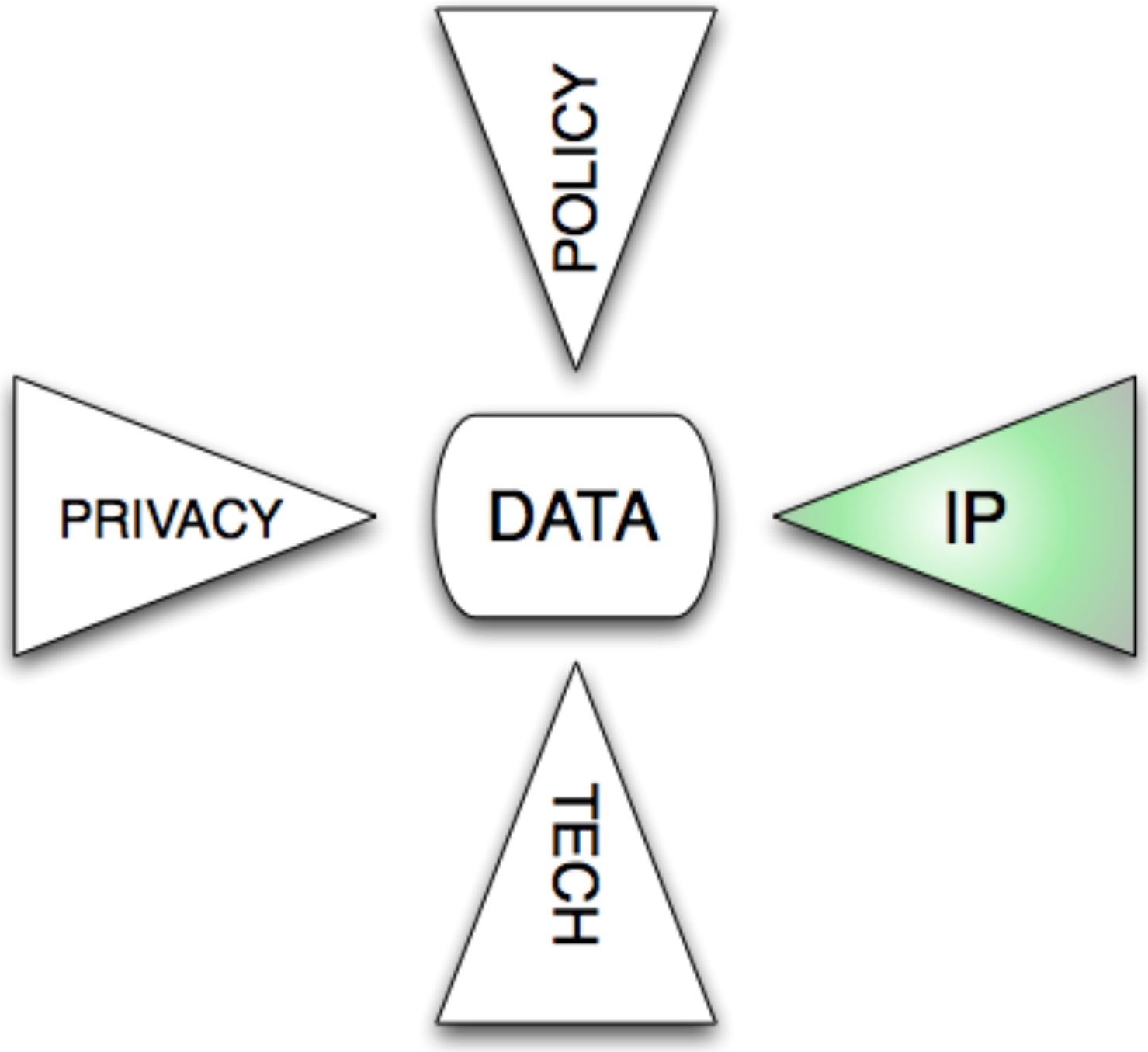


3. how might we “score” generativity?

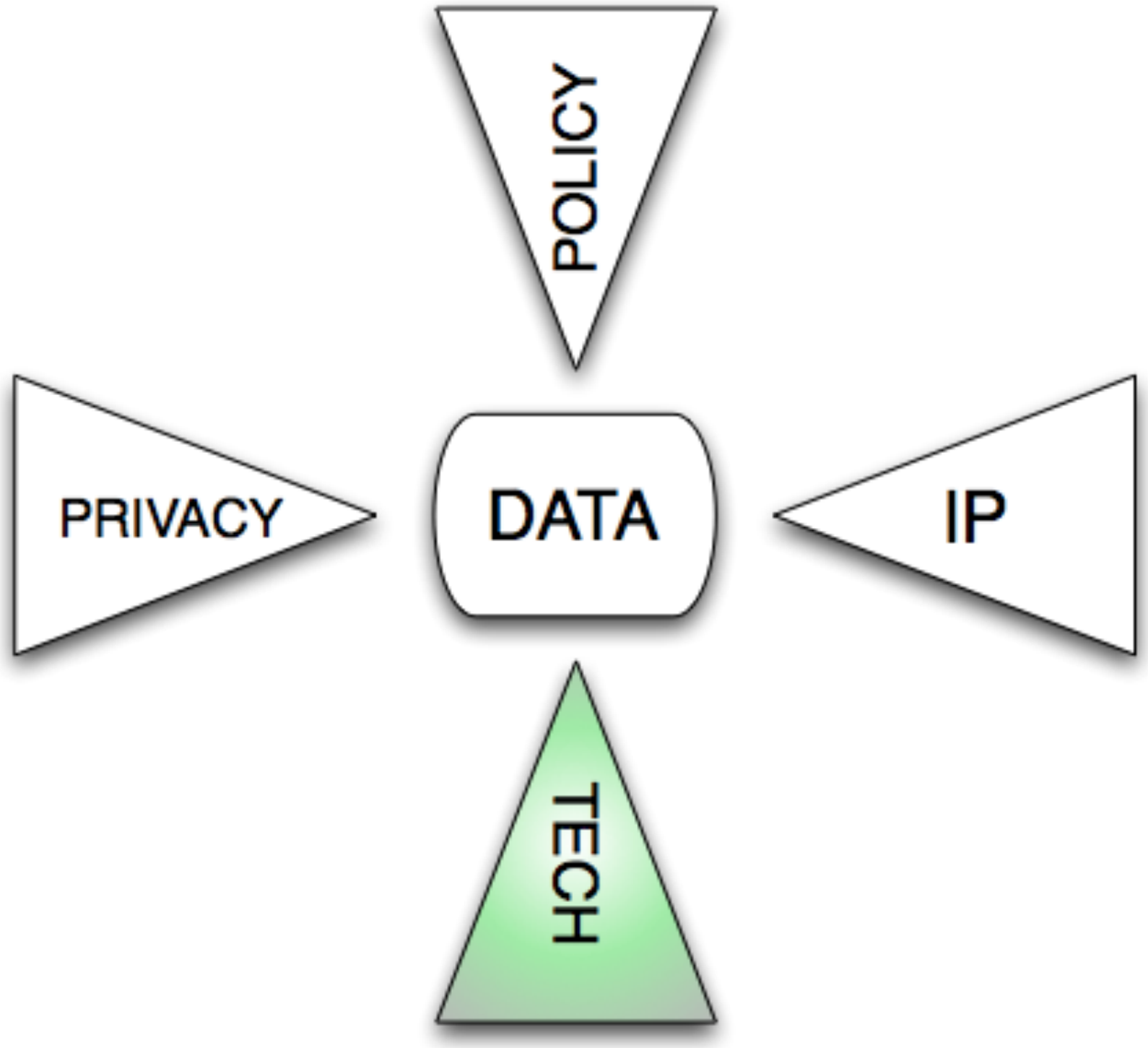

accessibility
adaptability
ease
of mastery
leverage

# what things regulate?

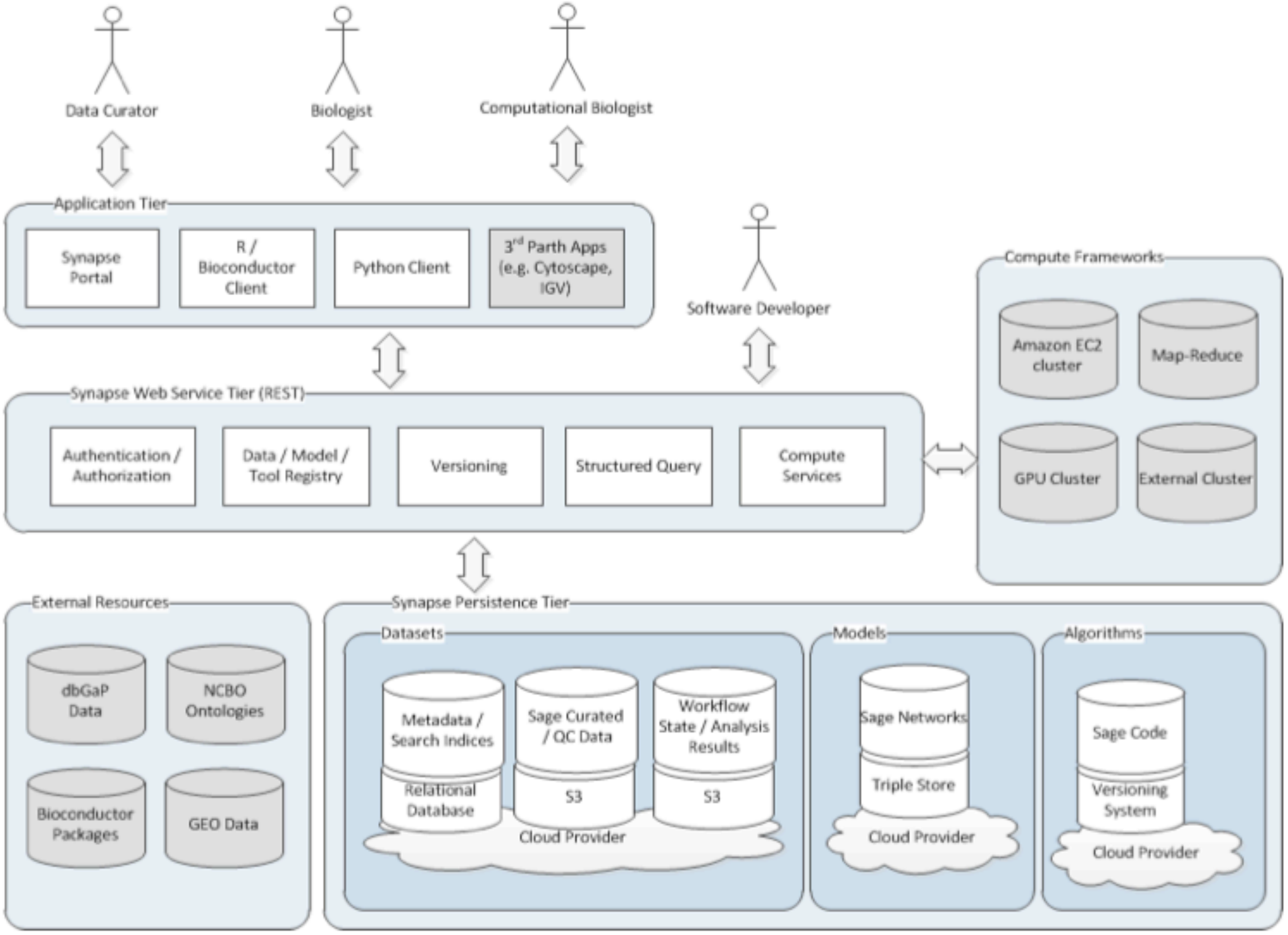


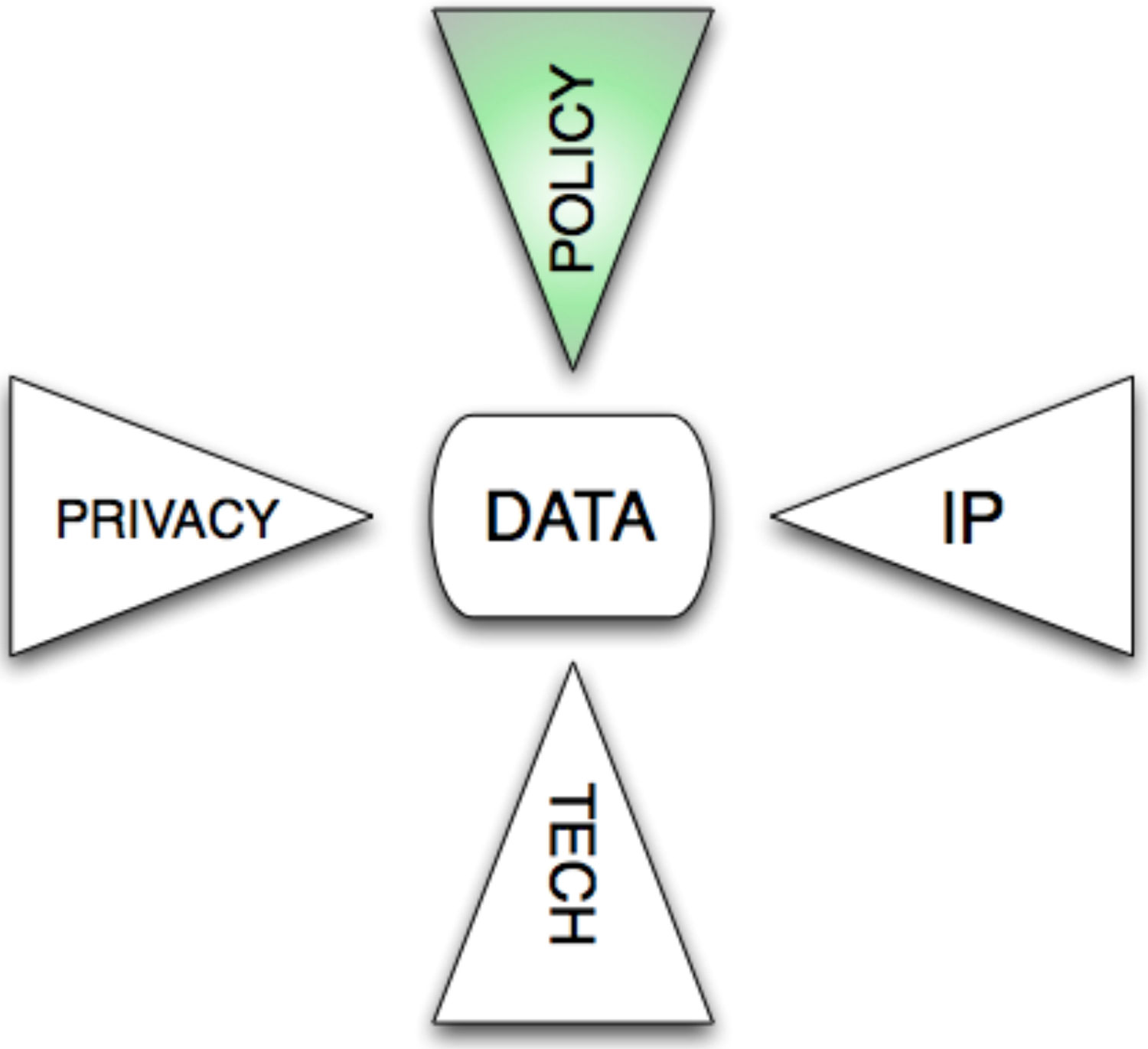

POLICY
PRIVACY
DATA
IP
TECH

copyright
database rights
patents
trade secrets
contracts

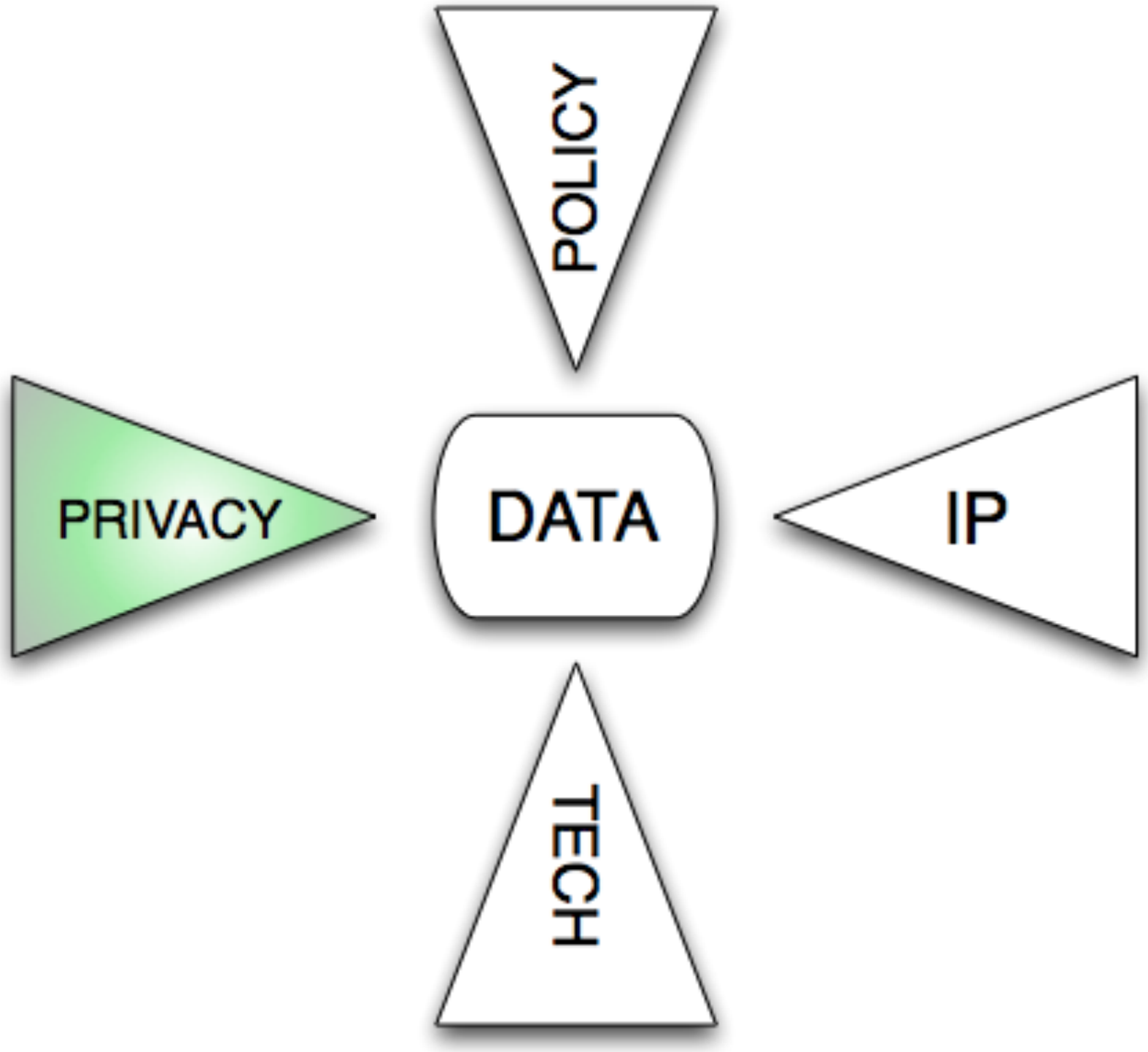

POLICY
PRIVACY
DATA
IP
TECH

ontology
API
format
OAuth
version control
provenance

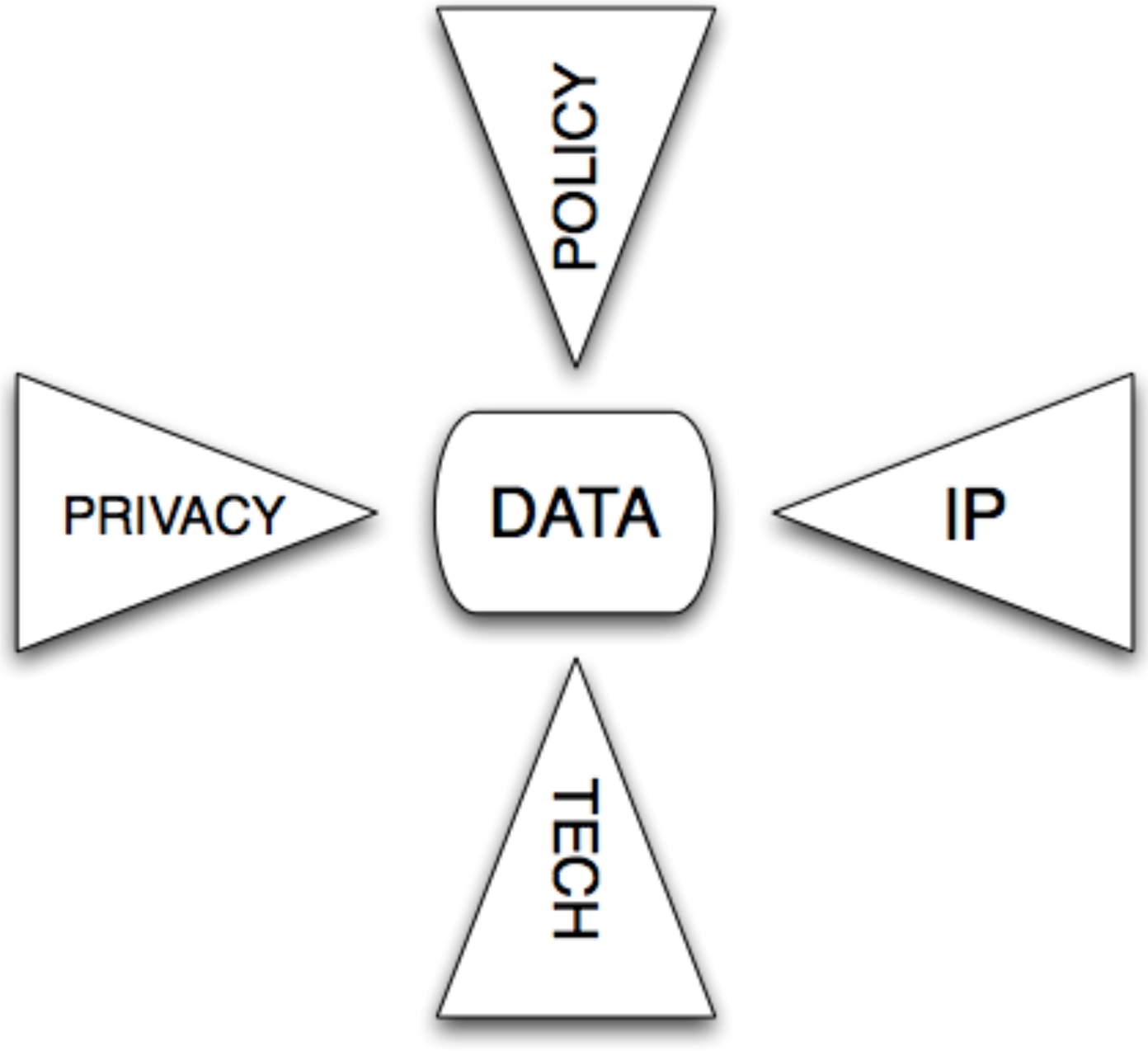

Data Curator
Biologist
Computational Biologist
Application Tier
Synapse Portal
R / Bioconductor Client
Python Client
3rd Parth Apps (e.g. Cytoscape, IGV)
Software Developer
Compute Frameworks
Amazon EC2 cluster
Map-Reduce
GPU Cluster
External Cluster
Synapse Web Service Tier (REST)
Authentication / Authorization
Data / Model / Tool Registry
Versioning
Structured Query
Compute Services
External Resources
dbGaP Data
NCBO Ontologies
Bioconductor Packages
GEO Data
Synapse Persistence Tier
Datasets
Metadata / Search Indices
Relational Database
Sage Curated / QC Data
S3
Workflow State / Analysis Results
S3
Cloud Provider
Models
Sage Networks
Triple Store
Cloud Provider
Algorithms
Sage Code
Versioning System
Cloud Provider


POLICY
PRIVACY
DATA
IP
TECH

open by default
safe harbors
estuary models
fund once, use many times


POLICY
PRIVACY
DATA
IP
TECH

tension between anonymity and utility
ease of re-identification
changing norms of privacy




what tools and approaches increase the generativity of data?

# The Licenses

**Attribution**
**CC BY**

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

View License Deed | View Legal Code

**Attribution-ShareAlike**
**CC BY-SA**

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

View License Deed | View Legal Code

**Attribution-NoDerivs**
**CC BY-ND**

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

View License Deed | View Legal Code

**Attribution-NonCommercial**
**CC BY-NC**

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

View License Deed | View Legal Code

**Attribution-NonCommercial-ShareAlike**
**CC BY-NC-SA**

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

View License Deed | View Legal Code

**Attribution-NonCommercial-NoDerivs**
**CC BY-NC-ND**

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

View License Deed | View Legal Code

# CC0 1.0 Universal (CC0 1.0)
# Public Domain Dedication

This is a human-readable summary of the **Legal Code (read the full text)**.

Disclaimer

## No Copyright

The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below.

## Other Information

- In no way are the patent or trademark rights of any person affected by CC0, nor are the rights that other persons may have in the work or in how the work is used, such as **publicity or privacy** rights.
- Unless expressly stated otherwise, the person who associated a work with this deed makes no warranties about the work, and disclaims liability for all uses of the work, to the fullest extent permitted by applicable law.
- When using or citing the work, you should not imply **endorsement** by the author or the affirmer.

# Consent to Research Getting Consent

LOGIN

**These are the rights you are granting to qualified researchers**

- ☐ Right to do research with my data
- ☐ Right to redistribute my data
- ☐ Right to publish the results of research from my data
- ☐ Right to commercialize products derived from research on my data

All boxes must be checked to move forward in the consent process

Next

GET INFORMED

- ✓ Welcome
- ✓ Researcher Terms of Use
- Grant Rights
- Watch Video
- Checkpoint
- Acknowledge Understanding

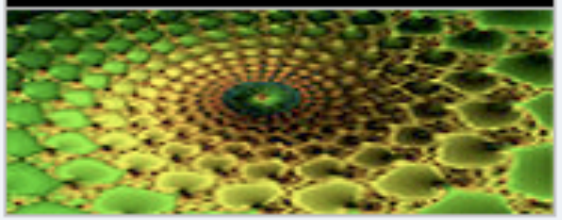
CONSENT TO RESEARCH

- Consent Form
- Generate ID

UPLOAD YOUR DATA

- User Profile
- Sign In
- Upload Data

**Office of Budget, Finance and Award Management (BFA)**



DIAS Home

CAAR Branch

Policy Office

Systems Office

View DIAS Staff

Search DIAS Staff

BFA Organization

Office of Budget, Finance, & Award Management

Budget Division

Division of Acquisition and Cooperative Support

# Dissemination and Sharing of Research Results

## NSF Data Sharing Policy

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

---

## NSF Data Management Plan Requirements

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

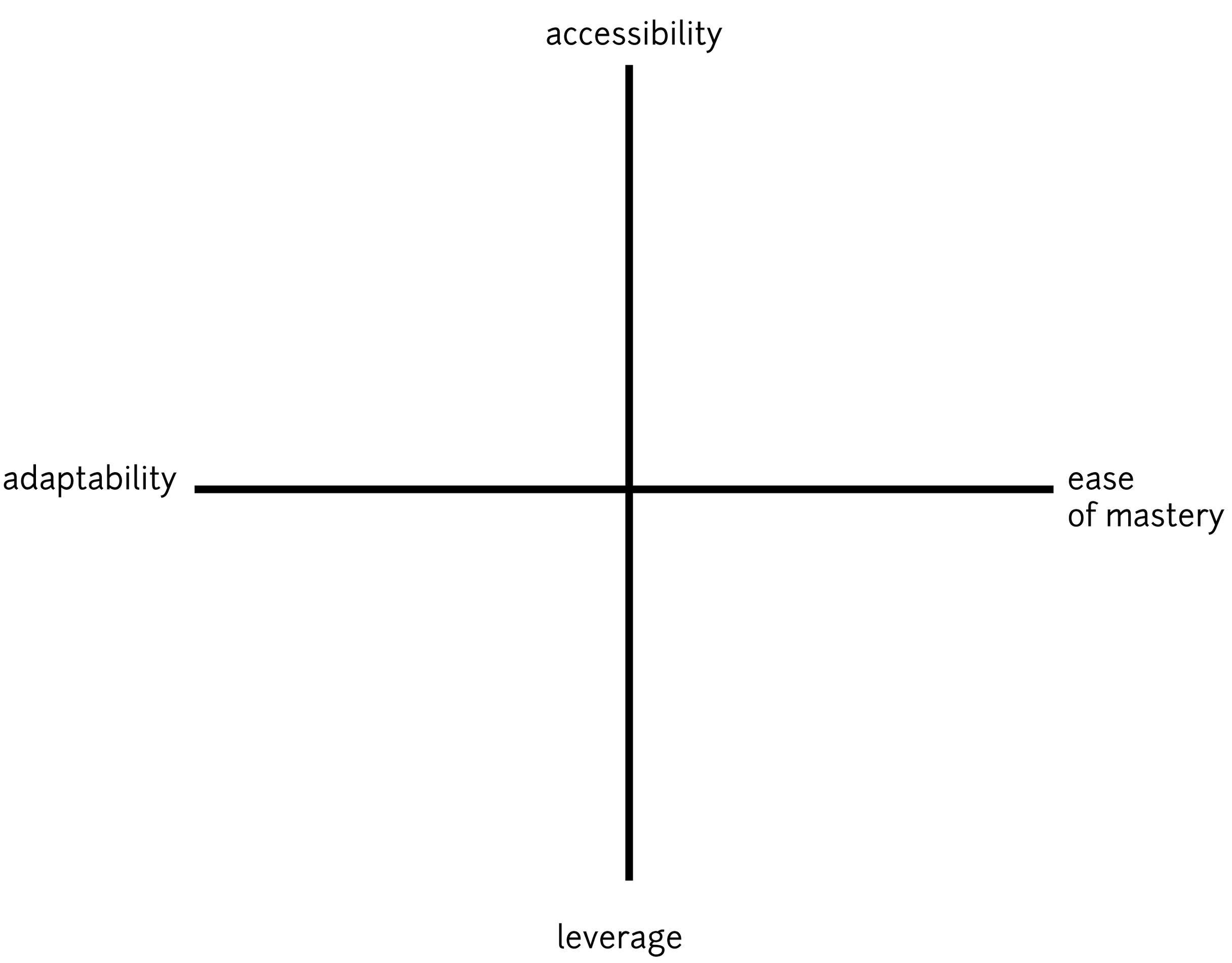
---

## Requirements by Directorate, Office, Division, Program, or other NSF Unit

Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in Grant Proposal

- No non-fuzzy understanding of why this is useful (“open data is good” is not enough) or of what interventions make open data more useful
- By more useful, we mean: more likely to result in unexpected repurposing of data

# 3. “scoring” generativity in open government data


accessibility
adaptability
ease
of mastery
leverage

# observations

- Adaptability correlates to accessibility
  - Legal and technical (both formats and raw downloads) accessibility makes data sets more obvious targets for adaptation
  - Not necessarily causal – some data sets are poorly accessible, but so important, that they are good targets anyway
  - Data.gov sets a good example
  - Accessibility is necessary, but not sufficient, for adaptability
- Accessibility is perhaps the easiest switch to flip if data is not of immediate economic value

# observations

- Adaptability correlates to leverage
  - Easier to do something unexpected if data is adaptable
  - Good annotations / metadata increase adaptability and leverage at the same time
  - Not necessarily causal
  - Domain expertise often needed to make leverage out of data, but is often not the domain expertise used to create the data or annotate it
  - Adaptability and accessibility necessary, but not sufficient, for leverage
- Adaptability is costly compared to accessibility due to need for human engagement in creation of metadata and annotation – and thus is more rare than accessibility.

# observations

- Pursuit of ease of mastery can ignore other three
  - Focus on user experience assumes naïve user, often doesn't expose raw data or address licensing or annotations
  - "Accessibility" when converted to "accessible to the layman" doesn't count in generativity terms for acceess
  - Ease of mastery often focuses on obscuring the complexity of the data, which can correlate negatively to adaptability or leverage
- However, a focus on ease of mastery + accessibility may be the sweet spot – click here for laypeople and get UX, click here if a data scientist and get the raw data

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

# Western Wind Resources Dataset

This interface gives access to the 30,000+ sites that were modeled by 3TIER as part of the Western Wind and Solar Integration Study.

The data can be accessed in two ways:

- Use interactive map to zoom in and click on a turbine
- Choosing a Station ID from the metadata file and typing it into the form below

Enter Station Id:

go

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

NREL HOME

ABOUT NREL | ENERGY ANALYSIS | SCIENCE & TECHNOLOGY | TECHNOLOGY TRANSFER | TECHNOLOGY DEPLOYMENT | ENERGY SYSTEMS INTEGRATION

# Disclaimer

More Search Options
Site Map
SEARCH

Printable Version

The National Renewable Energy Laboratory (NREL) is operated for the U.S. Department of Energy by the Alliance for Sustainable Energy, LLC ("Alliance"). As such the following rules apply:

## Copyright Status

NREL-authored documents are sponsored by the U.S. Department of Energy under Contract DE-AC36-08GO28308. Accordingly, with respect to such documents, the U.S. Government and others acting on its behalf retain a paid-up nonexclusive, irrevocable world-wide license to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. Use of documents available from or referenced by this server may be subject to U.S. and foreign Copyright Laws.

## Disclaimer of Endorsement

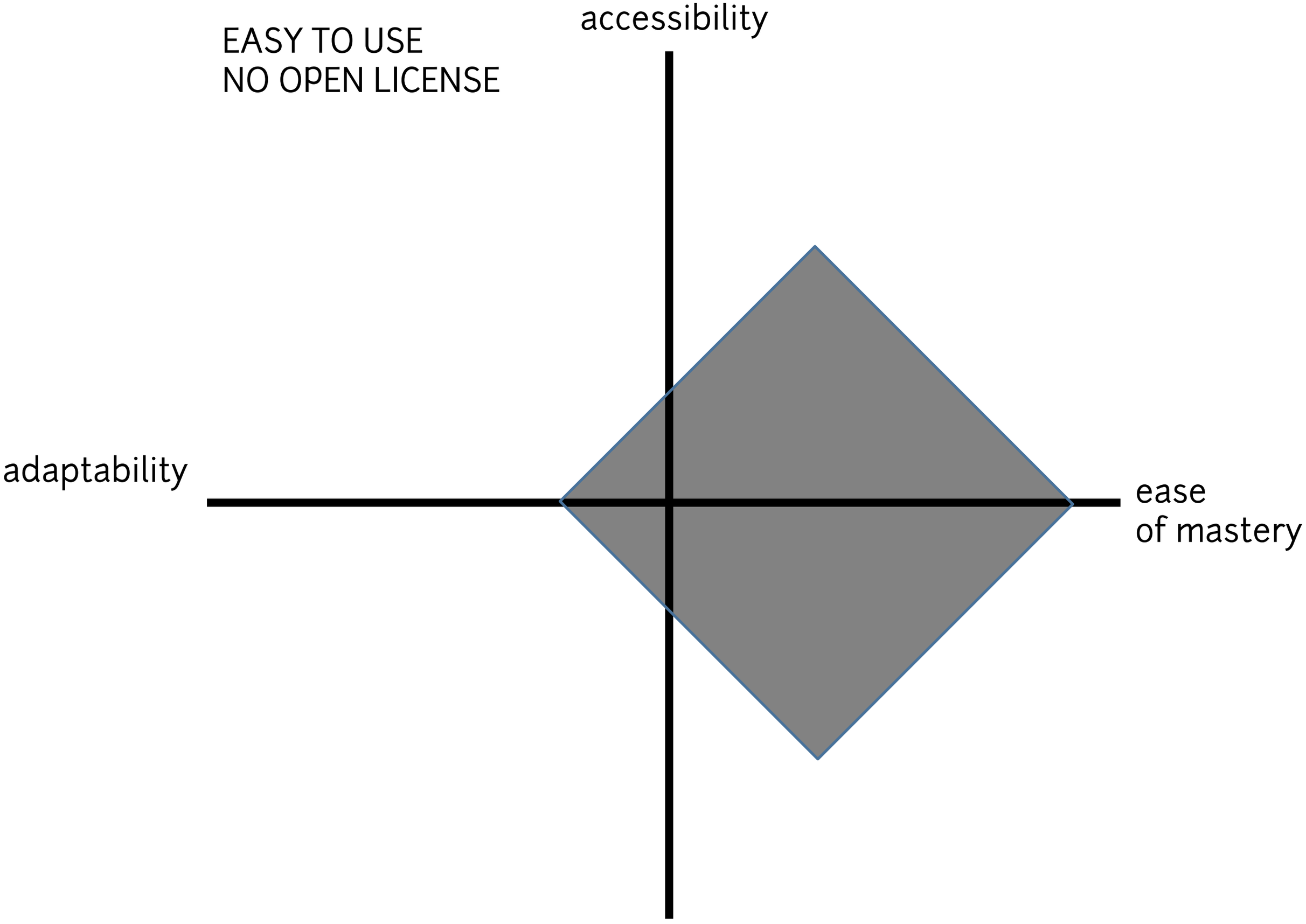
Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or Alliance . The views and opinions of authors expressed in the available or referenced documents do not necessarily state or reflect those of the United States Government or Alliance.

## Disclaimer of Liability

WITH RESPECT TO INFORMATION FROM THIS SERVER, NEITHER THE UNITED STATES GOVERNMENT NOR ALLIANCE, NOR ANY OF THEIR EMPLOYEES, MAKES ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING THE WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, OR ASSUMES ANY LEGAL LIABILITY OR RESPONSIBILITY FOR THE ACCURACY, COMPLETENESS, OR USEFULNESS OF ANY SUCH INFORMATION, OR OF ANY APPARATUS, PRODUCT, OR PROCESS DISCLOSED, OR REPRESENTS THAT ITS USE WOULD NOT INFRINGE PRIVATELY OWNED RIGHTS.

Printable Version

EASY TO USE
NO OPEN LICENSE



## NREL/3TIER - Western Wind Resources Dataset

Each turbine icon represents a site consisting of ten 3 MW Vestas V90 turbines (for a total of 30 MW).

The 30,000+ sites were selected from over 1.2 million points based on pre-existing sites; transmission corridors and RE zones; load correlation and wind power density.

This data was modeled by 3TIER using the Weather Research & Forecasting (WRF) model to downscale the NCEP/NCAR reanalysis data.

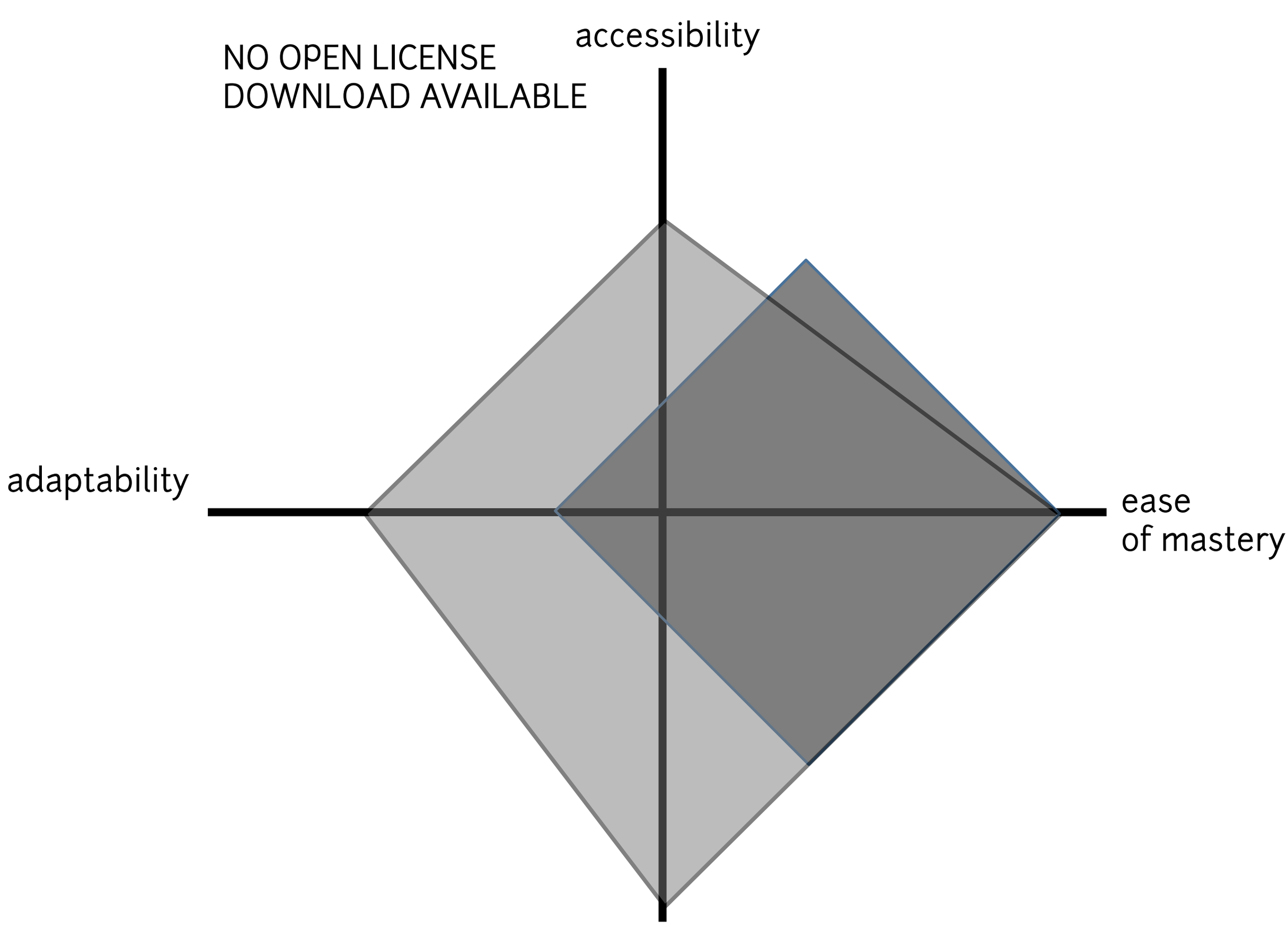
Close

Paste Clear

A1 SiteID

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | SiteID | Latitude | Longitude | Power Dens | SCORE-lite C | Wind Speed | State Code | Model Elevation [ |
| 2 | 1 | 31.192 | -102.242 | 413.925 | 30.823 | 7.741 | TX | 849.5 |
| 3 | 2 | 31.192 | -102.225 | 419.46 | 31.227 | 7.798 | TX | 859.4 |
| 4 | 3 | 31.192 | -102.208 | 429.712 | 32.042 | 7.911 | TX | 873.6 |
| 5 | 4 | 31.208 | -102.258 | 385.712 | 29.985 | 7.668 | TX | 902.7 |
| 6 | 5 | 31.208 | -102.242 | 375.495 | 29.305 | 7.609 | TX | 908.1 |
| 7 | 6 | 31.208 | -102.225 | 378.675 | 29.539 | 7.655 | TX | 915.9 |
| 8 | 7 | 31.208 | -102.208 | 383.684 | 30.01 | 7.723 | TX | 922.4 |
| 9 | 8 | 31.225 | -102.242 | 447.048 | 34.466 | 8.145 | TX | 947.9 |
| 10 | 9 | 31.225 | -102.225 | 441.857 | 34.324 | 8.146 | TX | 953.1 |
| 11 | 10 | 31.225 | -102.208 | 437.837 | 34.109 | 8.131 | TX | 948.7 |
| 12 | 11 | 31.275 | -104.542 | 425.601 | 30.468 | 7.738 | TX | 1344.4 |
| 13 | 12 | 31.275 | -104.525 | 453.123 | 31.794 | 7.907 | TX | 1363.2 |
| 14 | 13 | 31.292 | -104.525 | 440.34 | 30.684 | 7.797 | TX | 1361.8 |
| 15 | 14 | 31.292 | -104.458 | 562.009 | 33.463 | 8.075 | TX | 1284.2 |
| 16 | 15 | 31.292 | -104.442 | 540.765 | 32.127 | 7.926 | TX | 1277.7 |
| 17 | 16 | 31.308 | -104.508 | 477.308 | 31.735 | 7.911 | TX | 1344.1 |
| 18 | 17 | 31.308 | -104.492 | 524.97 | 33.154 | 8.06 | TX | 1320.3 |
| 19 | 18 | 31.308 | -104.475 | 516.722 | 32.106 | 7.947 | TX | 1309.2 |
| 20 | 19 | 31.325 | -104.525 | 455.937 | 30.635 | 7.843 | TX | 1370.5 |
| 21 | 20 | 31.325 | -104.458 | 523.132 | 31.632 | 7.966 | TX | 1320.3 |
| 22 | 21 | 31.342 | -104.542 | 423.342 | 28.901 | 7.676 | TX | 1398.2 |
| 23 | 22 | 31.342 | -104.458 | 486.605 | 30.157 | 7.812 | TX | 1341 |
| 24 | 23 | 31.342 | -104.375 | 455.77 | 28.751 | 7.574 | TX | 1330.3 |
| 25 | 24 | 31.342 | -104.358 | 454.683 | 28.451 | 7.485 | TX | 1298.3 |
| 26 | 25 | 31.358 | -104.508 | 482.053 | 30.973 | 7.905 | TX | 1390.7 |
| 27 | 26 | 31.358 | -104.442 | 502.848 | 30.961 | 7.913 | TX | 1347.8 |
| 28 | 27 | 31.358 | -104.425 | 479.688 | 30.067 | 7.805 | TX | 1353.6 |
| 29 | 28 | 31.358 | -104.408 | 487.624 | 30.527 | 7.854 | TX | 1358.1 |
| 30 | 29 | 31.358 | -104.392 | 459.859 | 29.323 | 7.674 | TX | 1349 |
| 31 | 30 | 31.358 | -104.375 | 467.813 | 29.509 | 7.662 | TX | 1330 |

NO OPEN LICENSE
DOWNLOAD AVAILABLE



## About the Weather Research & Forecasting Model

The Weather Research and Forecasting (WRF) Model is a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. It features multiple dynamical cores, a 3-dimensional variational (3DVAR) data assimilation system, and a software architecture allowing for computational parallelism and system extensibility. WRF is suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers.

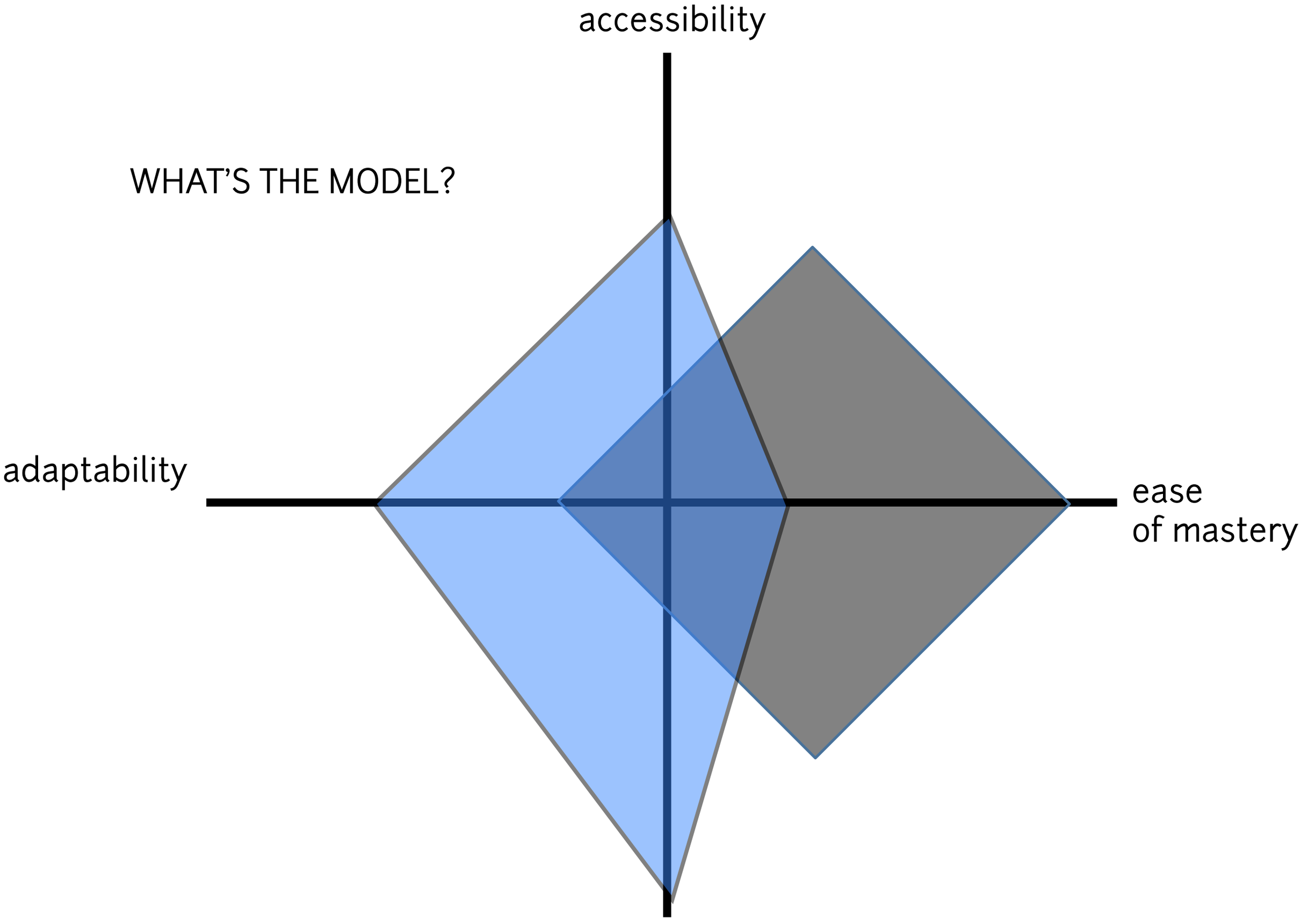
The effort to develop WRF has been a collaborative partnership, principally among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (the National Centers for Environmental Prediction (NCEP) and the Forecast Systems Laboratory (FSL), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA). WRF allows researchers the ability to conduct simulations reflecting either real data or idealized configurations. WRF provides operational forecasting a model that is flexible and efficient computationally, while offering the advances in physics, numerics, and data assimilation contributed by the research community.

# This is modeled (not measured) wind data

Western US:

- 3TIER generates historical wind data by running a Numerical Weather Prediction Model using physical conservation equations that 'recreate the weather' for 2004-6.
- They sampled the weather at a 1 arc-minute (~2km) spatial and 10 minute temporal resolution, and at 5 hub heights (10, 20, 50, 100, 200m).
- Based on a limited number of actual tower measurements for that time period, they did a sophisticated adjustment of (MOS-corrected) the model so that the data more accurately reflects actual wind speed measurements.
- Validation reports comparing the model wind speed results to actual data will be produced by 3TIER when their modeling is complete.
- 3TIER used SCORE-lite to convert wind speed to power output, assuming a Vestas V90 3 MW turbine at 100m hub height. Each grid point can hold 10 turbines or a 30 MW wind plant. The SCORE-lite process applies a probability distribution function to the manufacturer's power curve to replicate actual wind farm output.
- 3TIER produced hourly forecasts for day-ahead wind power output for the 30,000 selected sites.

WHAT'S THE MODEL?



# 4. “private” investments in generativity in open government data

Home

# TCGA Data Portal Overview

We provide 3 ways to download data: The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high-throughput sequencing analysis of the tumor genomes.

**The TCGA Data Portal does not host lower levels of sequence data. NCI's Cancer Genomics Hub (CGHub) is the new secure repository for storing, cataloging, and accessing BAM files and metadata for sequencing data. New users must still apply for authorized access through NCBI's Database of Genotypes and Phenotypes (dbGaP).**

Download Data

Choose from three ways to download data

| Available Cancer Types | # Cases Shipped by BCR | # Cases with Data* | Date Last Updated (mm/dd/yy) |
|---|---|---|---|
| Acute Myeloid Leukemia [LAML] | 200 | 200 | 07/16/13 |
| Adrenocortical carcinoma [ACC] | 80 | 80 | 07/22/13 |

## Announcements

### 06/13/2013 - DCC Software Released

The software release scheduled for today is complete and the TCGA Data Portal has been returned to normal operation. As part of this release, a new version of the TCGA Archive Validator has been provided and we strongly suggest that data submitting centers download and use this new version. The TCGA Archive Validator can be found on the TCGA Wiki here.

A complete list of the items addressed in this release can be found on the TCGA Wiki here and for those with JIRA access the tickets covered in this release can be found on the wiki here.

If you have any questions or concerns about this release, contact tcga-dcc-binf-l@list.nih.gov.

### 05/21/2013 - DCC Software Released

# Download Data

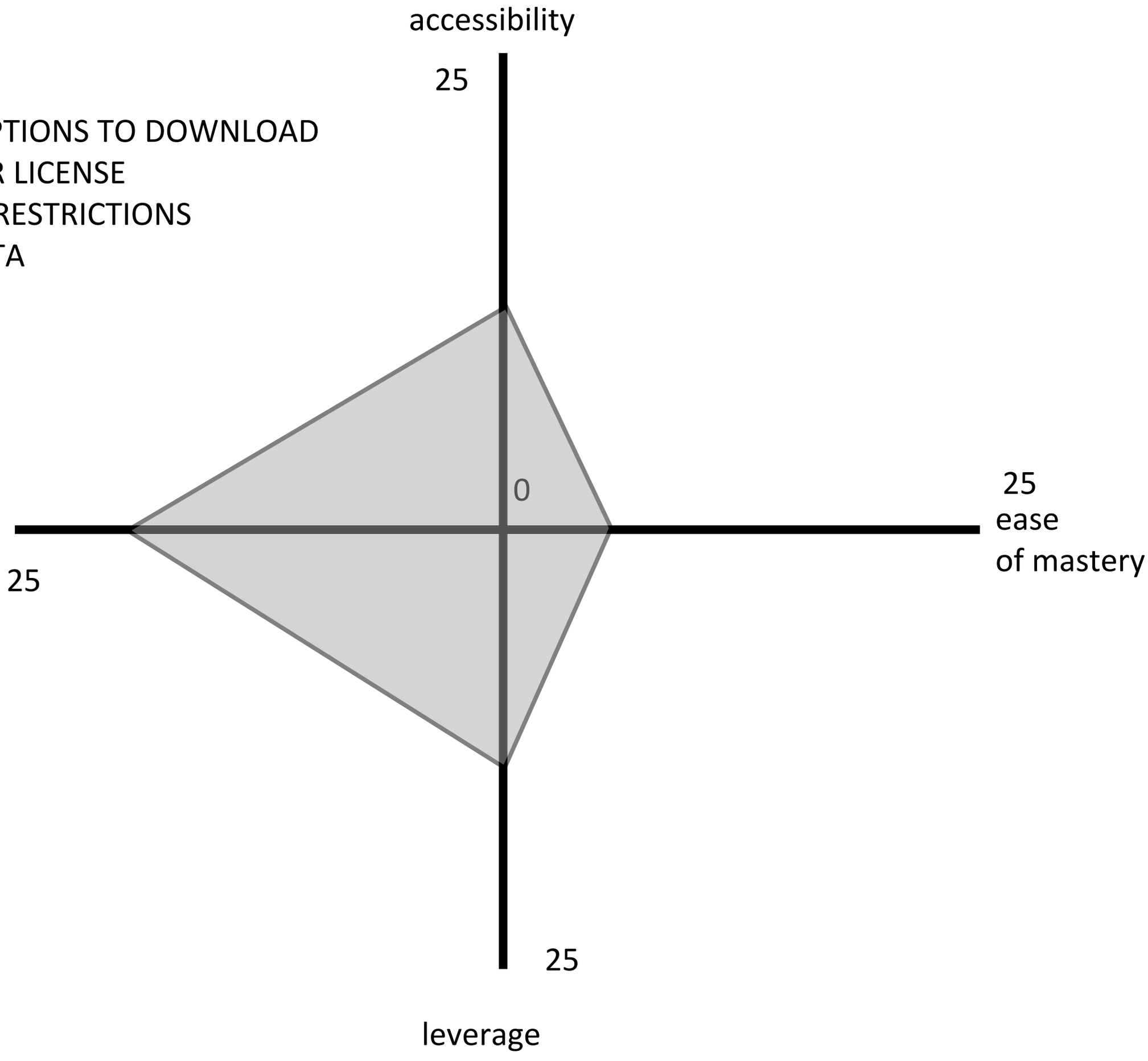
**The TCGA Data Portal does not host lower levels of sequence data. NCI's Cancer Genomics Hub (CGHub) is the new secure repository for storing, cataloging, and accessing BAM files and metadata for sequencing data. New users must still apply for authorized access through NCBI's Database of Genotypes and Phenotypes (dbGaP).**

**IMPORTANT:** Data downloaders are urged to use the data annotation search interface (https://tcga-data.nci.nih.gov/annotations/) to query the case, sample, and aliquot identifiers in their download to obtain the latest information associated with their data.

We provide 3 ways to download data:

| Method | What it offers | When to use it |
|---|---|---|
| Data Matrix | Select and download subsets of data by center, platform and data types.<br><br>Includes: Level 1, 2 and 3 data<br><br>Access the FAQ | Use when:<br>▪ You want to download data as tab-delimited text<br>▪ You only want a subset of the data |
| Bulk Download | A form that helps you locate files in the data archives.<br><br>Includes: Level 1, 2, 3 and limited level 4 data | Use when:<br>▪ You want to download bulk datasets as provided by the research centers |
| **Access HTTP Directories**<br>▪ Open-access HTTP Directory<br>▪ Controlled-Access HTTP Directory | Direct access to the HTTP directories where the data archives are stored.<br><br>Includes: Level 1, 2, 3 and limited level 4 data.<br><br>Login is required for the Controlled-access HTTP Directory. See controlled-access requirements. | Use when:<br>▪ You know how to use HTTP directories and you prefer to find files yourself rather than use the Bulk Download form |

THREE OPTIONS TO DOWNLOAD
NO CLEAR LICENSE
PRIVACY RESTRICTIONS
METADATA



which patients have tumor and normal samples amidst a group of celfiles?

# SCR TCGA

Added by Brig Mecham, last edited by Brig Mecham on Jun 11, 2012 (view change)

The Cancer Genome Atlas (TCGA) is an effort to generate and distribute high-throughput molecular data from more than 20 distinct cancers. For each cancer a variety of different molecular features are profiled for tens to hundreds of individual tumors. This data is a vital resource in the on-going fight against cancer.

Through the Synapse Commons Repository (SCR) we provide access to the unrestricted TCGA data **available through their data portal**. We have also standardized the clinical phenotypes to our Synapse Ontology, simplifying integrating these data with other cancer data contributed to the SCR.

**Feedback**

Please provide comm
below.

## Crawling TCGA

The TCGA data is organized by cancer-type, platform or technology, and then batch. Our crawler creates a unique layer for each cancer-type, platform, and batch in Synapse. **Read this for more information.**

## Standardizing Clinical Phenotypes

Read here for details on accessing clinical phenotype data: **Handling Clinical Phenotype Data**

Tools

# SCR TCGA Clinical-Phenotypes

3 Added by Brig Mecham, last edited by Brig Mecham on May 04, 2012 (view change)

## Introduction

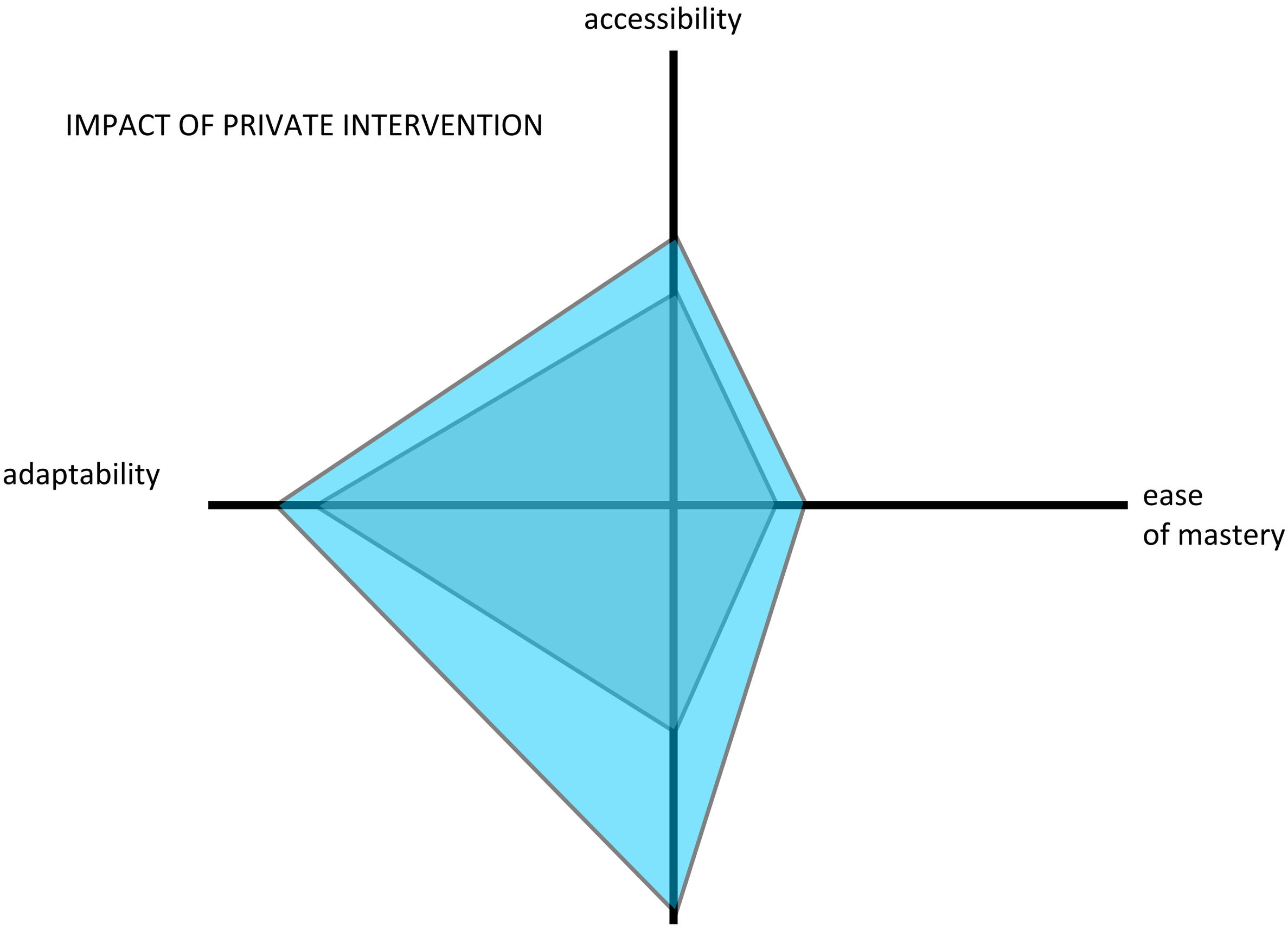
Our TCGA crawler simply builds layers for every file it finds on the TCGA ftp site. Each layers contain a location which allows users to work with the corresponding TCGA files through Synapse; for example, by using the getEntity or loadEntity commands in R. Before describing what we've done we'll briefly introduce the clinical information available from TCGA. First are the clinical files that contain information about the patients, their tumors, and the process used to generate data from any aliquot of biological material assayed by TCGA. Linking aliquots of genomic material to a specific data file is accomplished by merging the MAGE-tab files. Each technology used to process data from a given cancer has one or many MAGE-tab files. By merging these files we can get information about which patients DNA sample was hybridized to which Affy SNP 6.0 CEL file, as well the corresponding clinical variables about the patient (e.g. age, gender, stage), and some understanding of the experimental process used to generate the data (e.g. batches, collection centers, etc). Here we demonstrate how to access this information through the rSCR R library.

## handleTcgaClinical()

We developed an R function to merge all the information from a given cancer type into a single file. The following R command demonstrates how to use the R Synapse Client to download the clinical layers from Rectal Adenocarcinoma (READ) cancer.

```
> library(rSCR)
> ent <- loadEntity('syn220876')
```

IMPACT OF PRIVATE INTERVENTION



**The White House**

Office of the Press Secretary

E-Mail Tweet Share +

For Immediate Release

May 09, 2013

# Executive Order -- Making Open and Machine Readable the New Default for Government Information

EXECUTIVE ORDER

- - - - - - -

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT
FOR GOVERNMENT INFORMATION

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

open as in philosophy or
open as in methodology?

lots of implicit assumptions about data, scale, openness, value(s), justice, law, innovation.

free as in speech or free as in beer?

Paul McCartn
Play Fenway
Resume
T Stop
D-4

# Estimated Yearly Costs of Pet Ownership

**Stephen Zawistowski, Ph.D., Sr. VP Animal Sciences, ASPCA**

| Costs | Notes | Small Dog | Medium Dog | Large Dog | Cat | Rabbit | Guinea Pig | Small Mammal |
|---|---|---|---|---|---|---|---|---|
| **Annual Costs** | | | | | | | | |
| Food | 1 | $150.00 | $250.00 | $350.00 | $120.00 | $110.00 | $75.00 | $50.00 |
| Medical | 2 | $150.00 | $175.00 | $200.00 | $150.00 | $125.00 | $50.00 | |
| Litter | 3 | | | | $150.00 | $400.00 | $400.00 | $220.00 |
| Toys/Treats | | $50.00 | $60.00 | $70.00 | $50.00 | $25.00 | $25.00 | $10.00 |
| License | | $15.00 | $15.00 | $15.00 | | | | |
| Misc. | | $35.00 | $45.00 | $65.00 | $30.00 | $15.00 | $15.00 | $15.00 |
| Annual Total | | $400.00 | $545.00 | $700.00 | $500.00 | $675.00 | $565.00 | $295.00 |
| **Capital Costs** | | | | | | | | |
| Spay/neuter | | $75.00 | $100.00 | $125.00 | $75.00 | $75.00 | | |
| Collar/Leash | | $25.00 | $30.00 | $35.00 | $10.00 | | | |
| Litter box | | | | | $25.00 | $25.00 | | |
| Cage | | | | | | $80.00 | $80.00 | $35.00 |
| Carrier | | $30.00 | $50.00 | $80.00 | $30.00 | $30.00 | | |
| Crate | | $50.00 | $90.00 | $160.00 | | | | |
| Aquarium eqpt. | 4 | | | | | | | |
| Capital Total | | $180.00 | $270.00 | $400.00 | $140.00 | $210.00 | $80.00 | $35.00 |
| **Special Costs** | | | | | | | | |
| Long Hair Groom | | $200.00 | $300.00 | $400.00 | | | | |

john.wilbanks@sagebase.gov
wilbanks@nitrd.gov
@wilbanks